

Statistics

Lecture 2



Feb 19-8:47 AM

SG 3 & 4

Making a Freq. Table to organize data

class limits	class BNDRS	class MP	class F.	Cum. F.	Rel. F.	%F

Draw some statistical graphs

1) Bar chart

4) Freq. Poly gon

2) Histogram

5) Pie chart

3) Ogive

6) Stem Plot

7) Box Plot

Jun 24-6:49 PM

I randomly Selected 20 students, here are their ages:

1) $n = 20$

2) Min = 18
Max = 50

3) Range = $50 - 18 = 32$
Max - Min = 32 ✓

4) Midrange = $\frac{50 + 18}{2} = 34$ ✓

5) Mode 28 & 32
Bimodal

I want to make a freq. table with 3 classes.

class width = $\frac{\text{Range}}{\# \text{ classes}}$

If decimal → Round-up
If whole → Add 1

CW = $\frac{\text{Range} = 32}{3} = 10.66 \dots$
CW = 11

Jun 24-6:54 PM

Min

Class limits	class BNDs	Class MP	Class F	Cum. F	Rel. F	% F
18 - 28	17.5 - 28.5	23	8	8	.40	40%
29 - 39	28.5 - 39.5	34	7	15	.35	35%
40 - 50	39.5 - 50.5	45	5	20	.25	25%

class MP = $\frac{\text{+ class limits}}{2}$

Rel. F = $\frac{f}{n} = \frac{?}{20}$, % F = Rel. F (100)

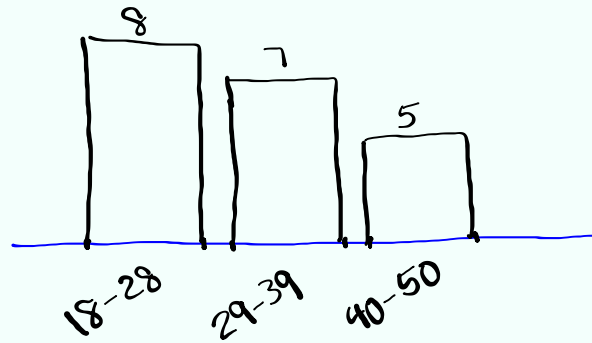
Jun 24-7:02 PM

Bar chart

- class limits

- class F

Has gaps

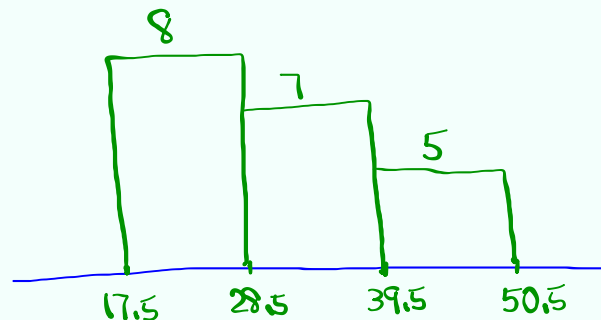


Histogram

- class BNDERS

- class F

NO GAPS



Jun 25-4:46 PM

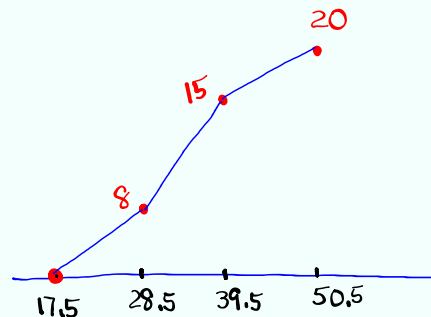
Ogive

- class BNDERS

- Cum. F

- Starts at 0

- Increasing
(goes up)



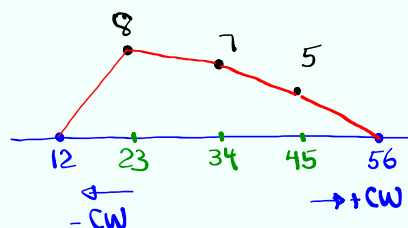
Freq. Polygon

- class MP

- Additional MP
one on each side

- class F

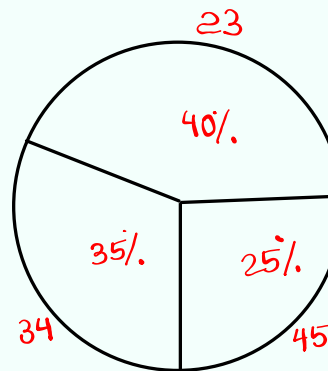
- Starts & stops
at 0 level.



Jun 25-4:50 PM

Pie chart

- Circle
- % F (Size of slice)
- class Mp (Name of slice)



STEM Plot

Data must be Sorted

18	20	20	24	26	1	8
28	28	28	30	32	2	0046888
32	32	35	37	38	3	0222578
40	42	45	50	60	4	025
					5	00

Jun 25-4:56 PM

I randomly selected 25 exams and here are the Scores.

55 58 63 68 69
70 72 75 75 75
78 80 80 82 86
86 86 86 90 92
95 98 100 100 100

- 1) $n = 25$
- 2) Min. = 55, Max = 100
- 3) Range = Max - Min
= 45
- 4) Midrange = $\frac{\text{Max} + \text{Min}}{2}$
= 77.5

5) Mode 86

6) Find class width if we wish to have 3

If decimal \Rightarrow Round up classes.

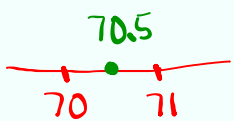
$$CW = \frac{\text{Range}}{\# \text{ classes}} = \frac{45}{3} = 15$$

If whole \Rightarrow Add 1

$$CW = 16$$

Jun 25-5:03 PM

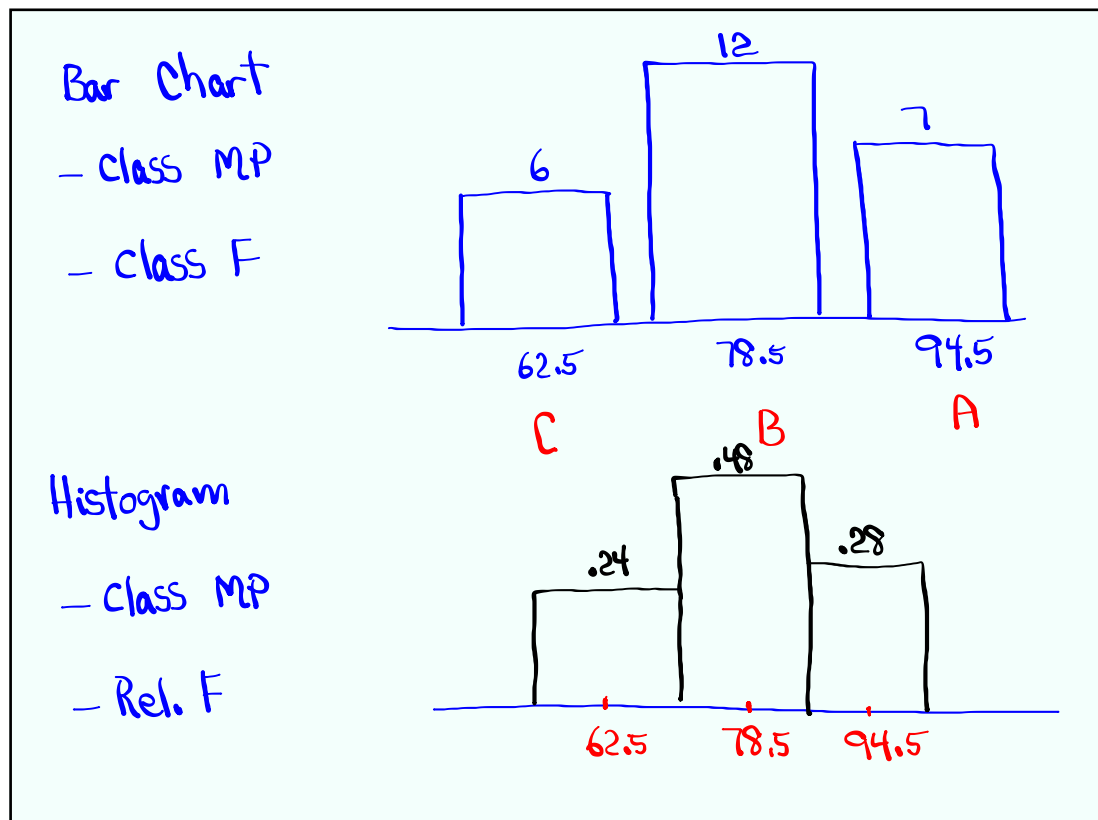
Class limits	class BNDs	Class MP	Class F	Cum. F	Rel. F	%F
55 - 70	54.5 - 70.5	62.5	6	6	.24	24%
71 - 86	70.5 - 86.5	78.5	12	18	.48	48%
87 - 102	86.5 - 102.5	94.5	7	25	.28	28%



$$\text{class MP} = \frac{\text{+class limits}}{2}$$

$$\text{Rel. F} = \frac{f}{n} = \frac{f}{25}$$

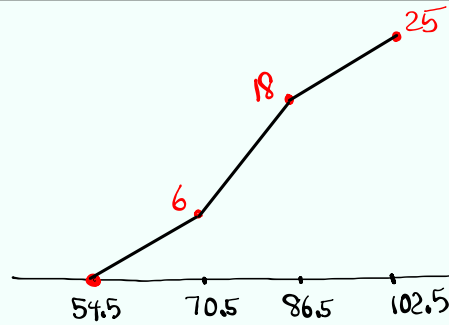
Jun 25-5:11 PM



Jun 25-5:22 PM

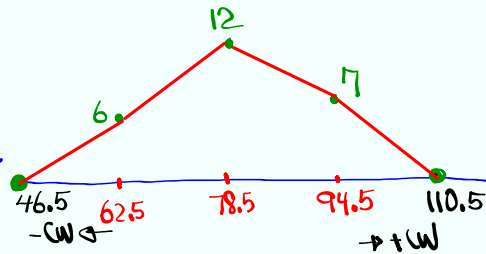
Ogive

- class BNDRS
- Cum. F
- start at 0 level
- Always increasing



Freq. Polygon

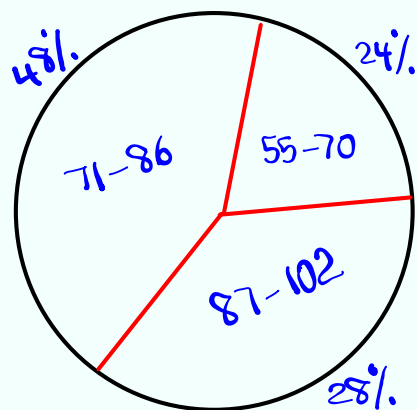
- class MP
- Additional MP one on each side
- start & stop at 0 level.



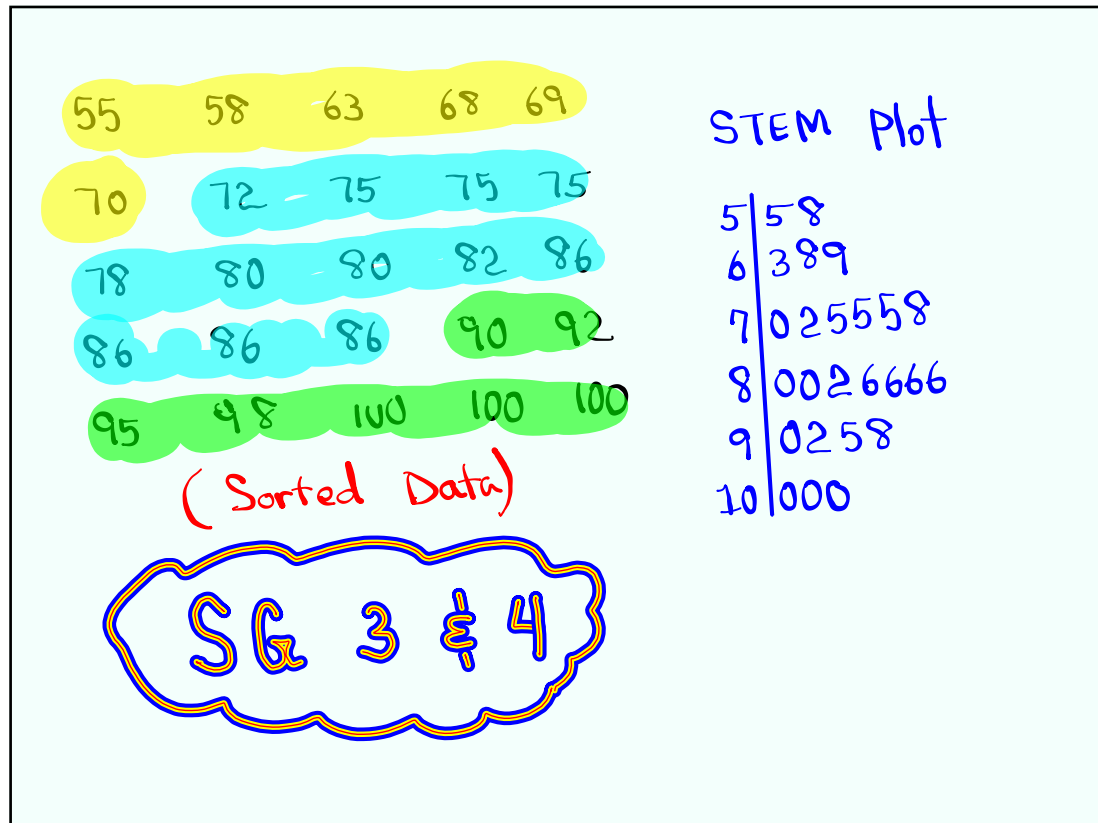
Jun 25-5:28 PM

Pie Chart

- Circle
- class limits (Name of Slice)
- % F (Size of Slice)



Jun 25-5:34 PM



Jun 25-5:39 PM

Basic Calculations in Statistic (SG 5-8)

$x \rightarrow$ Data element

$\sum x \rightarrow$ Sum of data elements

$n \rightarrow$ Sample Size

\bar{x} (x -bar) \rightarrow Sample Mean (Average)

ex: Consider the Sample below

2 3 3 5 7

1) $n=5$

2) Mode = 3

3) $\sum x = 2 + 3 + 3 + 5 + 7 = 20$

4) $\bar{x} = \frac{\sum x}{n} = \frac{20}{5} = 4$

Jun 25-6:00 PM

Consider the Sample below

1 3 4 6 7 9 10 13

1) $n = 8$

2) Range = $13 - 1 = 12$

3) Midrange = $\frac{13 + 1}{2} = 7$

4) Mode = None

5) $\sum x = 1 + 3 + 4 + \dots + 10 + 13 = 53$

6) $\bar{x} = \frac{\sum x}{n} = \frac{53}{8} = 6.625$

Round to

whole

1-dec.

2-dec.

Ans.

7

6.6

6.63

Jun 25-6:05 PM

$x \rightarrow$ Data element

$\sum x \rightarrow$ Sum of data elements

$\sum x^2 \rightarrow$ Square data element, then add.

$n \rightarrow$ Sample Size

$\bar{x} \rightarrow$ Sample Mean $\bar{x} = \frac{\sum x}{n}$

$S^2 \rightarrow$ Sample Variance

$$S^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

$$S^2 = \frac{n \sum x^2 - (\sum x)^2}{n(n - 1)}$$

ex: Consider the Sample below

1 3 3 3 5

$n = 5$

$\sum x = 1 + 3 + 3 + 3 + 5 = 15$

mode = 3

Median = 3

Range = 4

Midrange = 3

$\sum x^2 = 1^2 + 3^2 + 3^2 + 3^2 + 5^2 = 53$

$\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = 3$

$S^2 = \frac{n \sum x^2 - (\sum x)^2}{n(n - 1)} = \frac{5 \cdot 53 - 15^2}{5(5 - 1)} = \frac{265 - 225}{20} = \frac{40}{20} = 2$

Jun 25-6:13 PM

Consider the Sample below

1 3 3 5 5 7

1) $n = 6$

2) Mode = 3 & 5

3) Range = $7 - 1 = 6$

4) Midrange = $\frac{7+1}{2} = 4$

5) $\sum x = 24$

6) $\sum x^2 = 118$

7) $\bar{x} = \frac{\sum x}{n} = \frac{24}{6} = 4$

8) $S^2 = \frac{n \sum x^2 - (\sum x)^2}{n(n-1)} = \frac{6 \cdot 118 - 24^2}{6(6-1)} = \frac{132}{30} = 4.4$

$S^2 \rightarrow$ Sample Variance

$S^2 \geq 0$

$S \rightarrow$ Sample standard deviation

$S \geq 0$

$S = \sqrt{S^2}$

From last example

$S = \sqrt{4.4} \approx 2.098$

Jun 25-6:23 PM

Given

$n = 10$

$\sum x = 50$

$\sum x^2 = 294$

Min = 2

Max = 9

1) Range = Max - Min = 7

2) Midrange = $\frac{\text{Max} + \text{Min}}{2} = 5.5$

3) $\bar{x} = \frac{\sum x}{n} = \frac{50}{10} = 5$

4) $S^2 = \frac{n \sum x^2 - (\sum x)^2}{n(n-1)}$

5) $S = \sqrt{S^2}$

$= \sqrt{\frac{44}{9}} \approx 2.211$

$= \frac{10 \cdot 294 - 50^2}{10(10-1)}$

$= \frac{440}{9} = \frac{44}{9}$
 ≈ 4.889

Jun 25-6:32 PM

Given $n=8$ $\sum x=64$ $\sum x^2=512$

$$1) \bar{x} = \frac{\sum x}{n} = \frac{64}{8} = \boxed{8}$$

$$2) S^2 = \frac{n \sum x^2 - (\sum x)^2}{n(n-1)} = \frac{8 \cdot 512 - 64^2}{8(8-1)} = \frac{0}{56} = \boxed{0}$$

$$3) S = \sqrt{S^2} = \sqrt{0} = \boxed{0}$$

How to estimate S:

$$S \approx \frac{\text{Range}}{4} \quad \text{"the range rule-of-thumb"}$$

A sample has a min. 20 and max 100.

Estimate its standard deviation.

$$S \approx \frac{\text{Range}}{4} = \frac{100-20}{4} = \frac{80}{4} = \boxed{20}$$

Jun 25-6:39 PM

Assume data elements are sorted
from smallest to largest.

5-Number Summary

Min. Q_1 Median Q_3 Max

$Q_1 \rightarrow$ First Quartile

$Q_3 \rightarrow$ Third Quartile



25% below Q_1 , 75% above Q_1

75% below Q_3 , 25% above Q_3

50% below Median, 50% above Median

✓ Draw Box Plot



✓ IQR (Inter-Quartile-Range) = $Q_3 - Q_1$

✓ Upper Fence = $Q_3 + 1.5(IQR)$

✓ Lower Fence = $Q_1 - 1.5(IQR)$

Outliers

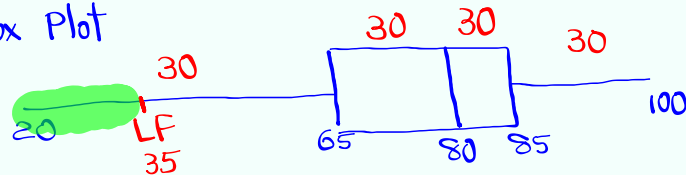
are below the lower fence or
above the upper fence

Jun 25-6:47 PM

I randomly selected 120 exams, the 5-Number Summary was $\rightarrow 120 \div 4 = 30$

20	65	80	85	100
Min	Q_1	MD	Q_3	Max

Box Plot



$$IQR = Q_3 - Q_1 = 85 - 65 = 20$$

$$\text{Upper Fence} = Q_3 + 1.5(IQR) = 85 + 1.5(20) = 115$$

$$\text{Lower Fence} = Q_1 - 1.5(IQR) = 65 - 1.5(20) = 35$$

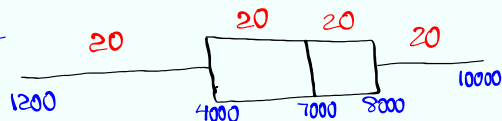
Possible outliers: 20 - 35

Jun 25-6:57 PM

I randomly selected 80 nurses, the following are the 5-Number Summary of their monthly Salaries. $80 \div 4 = 20$

	Q_1	MD	Q_3	Max
Min	4000	7000	8000	10,000

Box Plot



$$IQR = Q_3 - Q_1 = 4000$$

$$\text{Upper Fence} = Q_3 + 1.5(IQR) = 8000 + 1.5(4000) = 14000$$

$$\text{Lower Fence} = Q_1 - 1.5(IQR) = 4000 - 1.5(4000) = -2000$$

Possible outliers \rightarrow None

Jun 25-7:05 PM